```
---
title: "A2 - Matrices y vectores aleatorios"
author: "Josue Salvador Cano Martinez"
date: "2022-10-04"
output: html_document
---
```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
```

1) Matriz de datos X = [1 4 3, 6 2 6, 8 3 3] que consta de 3 observaciones (filas) y 3 variables (c)

```{r}
library(MVN)
X = matrix(c(1,4,3,6,2,6,8,3,3), ncol=3)
X=t(X)
X
```
```

```
     [,1] [,2] [,3]
[1,]    1    4    3
[2,]    6    2    6
[3,]    8    3    3
```

```
## a) Media, varianza y covarianza de b'X y c'X

```{r}
bX = matrix(c(8,14,14), ncol=1)
cX = matrix(c(0,-8,5), ncol=1)

print('Media bX')
mean(bX)
print('Varianza bX')
var(bX)

print('Media cX')
mean(cX)
print('Varianza cX')
var(cX)

print('Covarianza')
cov(bX, cX)
```
```

```
[1] "Media bX"
[1] 12
[1] "Varianza bX"
     [,1]
[1,]   12
[1] "Media cX"
[1] -1
[1] "Varianza cX"
     [,1]
[1,]   43
[1] "Covarianza"
     [,1]
[1,]   -3
```

b) Determinante de S (matriz de var-covarianzas de X)

```{r}
S <- cov(X)
det(S)
```

```
[1] 0
```

c) Matriz de varianzas-covarianzas

```{r}
S
```

```
     [,1] [,2] [,3]
[1,] 13.0 -2.5  1.5
[2,] -2.5  1.0 -1.5
[3,]  1.5 -1.5  3.0
```

d) Valores y vectores propios de S

```{r}
print(eigen(S))
```

```
eigen() decomposition
$values
[1]  1.379150e+01  3.208497e+00 -7.859007e-17

$vectors
           [,1]       [,2]       [,3]
[1,]  0.9645458 -0.2295697 -0.1301889
[2,] -0.2076189 -0.3555080 -0.9113224
[3,]  0.1629288  0.9060418 -0.3905667
```

e) Argumentar si b'X y c'X son independientes o no.

No son independientes debido a que la covarianza es diferente de cero y los datos fueron extraídos a partir de multiplicación de matrices (no como resultado de una función que las relacione)

f) Varianza generalizada. Explicar el comportamiento de los datos de X basándose en los la variable generalizada, en los valores y vectores propios.

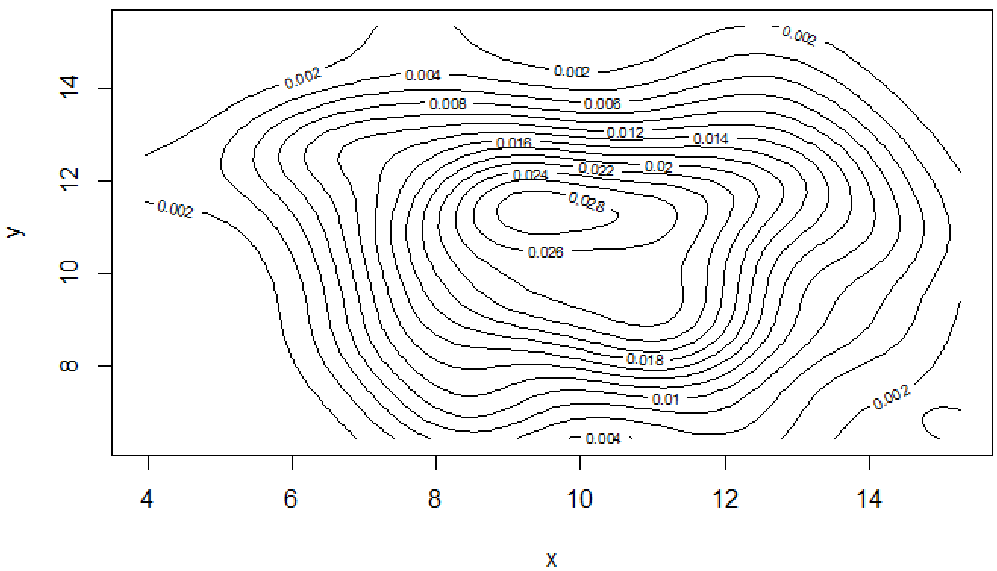
```{r}
det(cov(X))
```

```
[1] 0
```

Como la varianza generalizada es igual a 0 se infiere que existe una baja dispersión de los datos en x

```r
# 2.Explore los resultados del siguiente código y dé una interpretación:

```{r}
library(MVN)
x = rnorm(100, 10, 2)
y = rnorm(100, 10, 2)
datos = data.frame(x,y)
mvn(datos, mvnTest = "hz", multivariatePlot = "persp")
mvn(datos, mvnTest = "hz", multivariatePlot = "contour")
```
```



Tomando el valor p de la prueba de Henze-Zirkler se sabe que los datos tienen una distribución normal multivariante. La prueba de Anderson-Darling indica que ambas variables tienen una distribución normal.

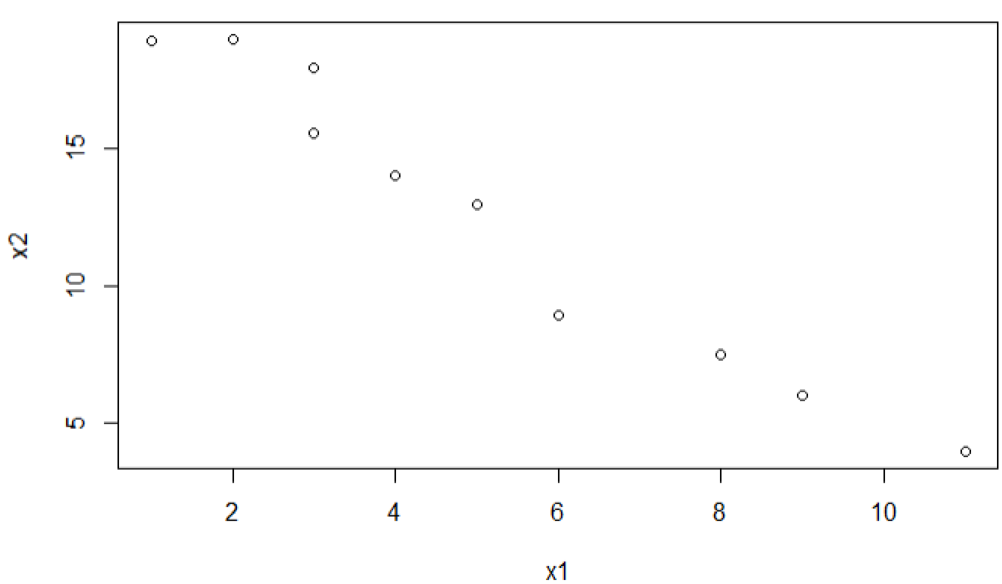
3.Un periódico matutino enumera los siguientes precios de autos usados para un compacto extranjero con edad medida en años y precio en venta medido en miles de dólares.

x1: 1,2,3,3,4,5,6,8,9,11

x2: 18.95, 19.00, 17.95, 15.54, 14.00, 12.95, 8.94, 7.49, 6.00, 3.99

a) Dagrama de dispersión

```{r}
x1 = c(1,2,3,3,4,5,6,8,9,11)
x2 = c(18.95,19.00,17.95,15.54,14.00,12.95,8.94,7.49,6.00,3.99)
plot(x1, x2)
```



b) signo de la covarianza muestral a partir del gráfico

Tomando en cuenta el diagrama de dispersión la covarianza muestral es negativa debido a que entre mayor es el valor de x1 el valor de x2 disminuye, es decir, el producto de (Xi-Xbarra) * (yi-ybarra) es negativo.

c) Calcular el cuadrado de las distancias estadísticas

```{r}
A = matrix(c(x1,x2), ncol = 2)
med = colMeans(A)
S = cov(A)
dist <- mahalanobis(A, med, S)
dist
```

```
[1] 1.8753045 2.0203262 2.9009088 0.7352659 0.3105192 0.0176162 3.7329012 0.8165401
[9] 1.3753379 4.2152799
```

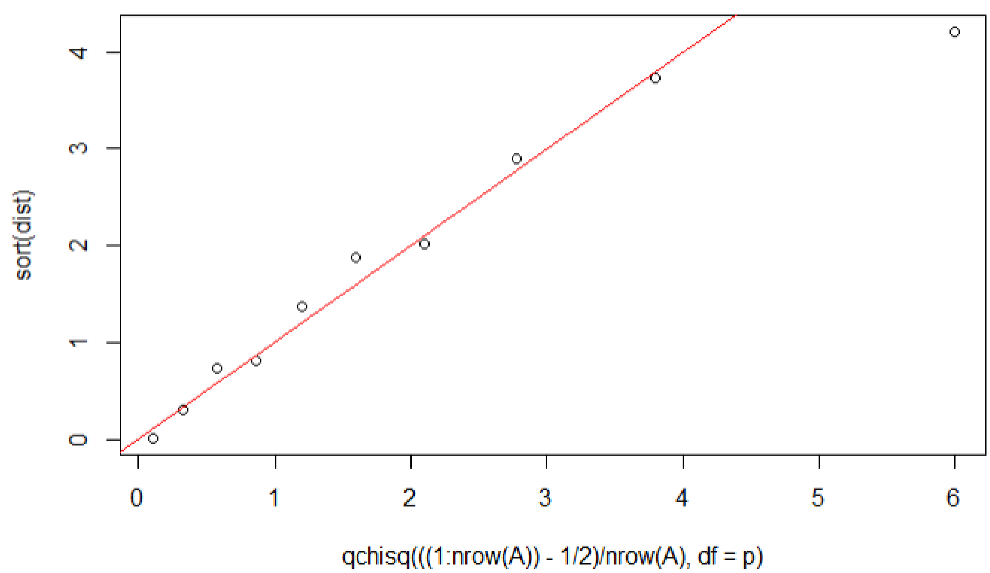
d) Proporción de las observaciones que caen dentro del contorno de probabilidad estimado del 50% de una distribución normal bivariada.

```{r}
p <- qchisq(((1:nrow(A)) - 1/2)/nrow(A),df=2)
p
```

```
[1] 0.1025866 0.3250379 0.5753641 0.8615658 1.1956740 1.5970154
[7] 2.0996442 2.7725887 3.7942400 5.9914645
```

e) Ordenar las distancias del inciso c y construir un diagrama chi-cuadrado

```{r}
p = 2
plot(qchisq(((1:nrow(A)) - 1/2)/nrow(A),df=p),sort(dist))
abline(a=0, b=1,col="red")
```



f) Dados los resultados anteriores, ¿serían argumentos para decir que los datos son aproximadamente normales bivariados?

Tomando en consideración el diagrama anterior se puede inferir que los datos son aproximadamente normales bivariados debido al comportamiento que presentan la mayoría de ellos al seguir el comportamiento de la gráfica (se encuentran sobre la línea roja), aún y cuando exista 1 valor atípico. Para confirmar resultaría útil un test de normalidad multivariada.